



AR in VR: assessing surgical augmented reality visualizations in a steerable virtual reality environment

Julian Hettig¹ · Sandy Engelhardt^{1,2} · Christian Hansen¹ · Gabriel Mistelbauer³

Received: 12 April 2018 / Accepted: 5 July 2018 / Published online: 24 July 2018
© CARS 2018

Abstract

Purpose Augmented reality (AR) has emerged as a promising approach to support surgeries; however, its application in real world scenarios is still very limited. Besides sophisticated registration tasks that need to be solved, surgical AR visualizations have not been studied in a standardized and comparative manner. To foster the development of future AR applications, a steerable framework is urgently needed to rapidly evaluate new visualization techniques, explore their individual parameter spaces and define relevant application scenarios.

Methods Inspired by its beneficial usage in the automotive industry, the underlying concept of virtual reality (VR) is capable of transforming complex real environments into controllable virtual ones. We present an interactive VR framework, called Augmented Visualization Box (AVB), in which visualizations for AR can be systematically investigated without explicitly performing an error-prone registration. As use case, a virtual laparoscopic scenario with anatomical surface models was created in a computer game engine. In a study with eleven surgeons, we analyzed this VR setting under different environmental factors and its applicability for a quantitative assessment of different AR overlay concepts.

Results According to the surgeons, the visual impression of the VR scene is mostly influenced by 2D surface details and lighting conditions. The AR evaluation shows that, depending on the visualization used and its capability to encode depth, 37% to 91% of the experts made wrong decisions, but were convinced of their correctness. These results show that surgeons have more confidence in their decisions, although they are wrong, when supported by AR visualizations.

Conclusion With AVB, intraoperative situations are realistically simulated to quantitatively benchmark current AR overlay methods. Successful surgical task execution in an AR system can only be facilitated if visualizations are customized toward the surgical task.

Keywords Surgical augmented reality · Virtual reality · Visualization

Introduction

With increasingly complex surgical procedures, intraoperative visualizations are becoming an important part of surgical environments. In this regard, augmented reality (AR) is a concept under wide investigation. It is expected to solve a major information deficit evident in current surgical procedures by

providing see-through vision of hidden objects under organ surfaces. However, AR systems are complex, as they mostly involve the integration of different technical steps. While most research has been conducted in the field of segmentation, registration and hardware development, little emphasis has been given to how anatomical AR visualizations can be improved. Yet, this step at the end of the processing pipeline must be considered as crucial given that the augmented information is ultimately perceived by the surgeon to make critical decisions, e.g., how to proceed to a tumor while avoiding delicate structures. As stated in the recent review paper by

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11548-018-1825-4>) contains supplementary material, which is available to authorized users.

✉ Sandy Engelhardt
sandy.engelhardt@isg.cs.uni-magdeburg.de

¹ Faculty of Computer Science & Research Campus STIMULATE, University of Magdeburg, Magdeburg, Germany

² Faculty of Computer Science, Mannheim University-of-Applied-Sciences, Mannheim, Germany

³ Faculty of Computer Science, University of Magdeburg, Magdeburg, Germany

Bernhardt et al. [2], most perceptual issues in surgical AR remain unsolved, especially when it comes to alleviating the prominent issue of depth perception.

Open and especially minimally invasive surgery (MIS) resembles a highly dynamic environment regarding motion, deformation and—in the case of AR visualizations—display and texture. Furthermore, most AR visualization techniques are only designed for a specific application scenario targeting a single organ and evaluated under rather ideal conditions, e.g., only statically on artificial medical phantom data [3,7] or animal cadavers [1,3,6,7]. Obstructive environmental influences, such as smoke, lens contamination or varying anatomical features that impair the perception of a visualization, are commonly not considered. Beyond that, two methods that are designated for the same surgical scenario can hardly be compared due to their individual evaluation strategies. Consequently, assessing the usefulness of the presented visualization in an objective manner is highly restricted.

In this paper, a steerable and interactive VR framework is presented, referred to as *Augmented Visualization Box (AVB)*, that offers the possibility to objectively and reproducibly evaluate AR visualization techniques in a self-contained manner. We introduce a dedicated system architecture and workflow in order to bypass the above-mentioned registration issues and foster the investigation of new visualization applications. We demonstrate the capabilities of AVB along a virtual laparoscopic liver scenario, configured with a variety of environmental influences.

In summary, the contributions of this paper are as follows:

- a steerable, interactive and parameterizable VR framework for the objective evaluation of AR visualization techniques in a surgical environment that avoids registration errors and
- a qualitative user study with eleven surgeons analyzing the applicability of the presented AVB framework.

Related work

Kersten-Oertel et al. [8] present a state of the art report on mixed reality visualization in image-guided surgery.

According to them, most visualizations in medical AR applications deal with simple representations of non-visible anatomical structures using transparency or color adjustments. More advanced techniques try to encode additional information, e.g., surgical planning data [6] or uncertainties [1]. Bichlmeier et al. [3] introduced an occlusion-based method called *virtual windowing* that employs transparency adjustments and clipping according to the patient's skin, location of instruments and line of sight of the observer. Lerotic et al. [9] proposed a non-photorealistic technique called *inverse realism* that integrates strong anatomical features of the AR-occluded surface and superimposes them

in turn on top of the augmentation. Hansen et al. [6] utilize *illustrative rendering* techniques for vasculature contours to encode perspective. Amir-Khalili et al. [1] also used *contour lines* to provide the surgeon with uncertainty information about the overlaid data. Nicolaou et al. [11] employed the concept of casting *virtual shadows* from instruments on the scene to improve depth perception. For an extensive overview, the reader is referred to the review by Bernhardt et al. [2]. The authors describe the individual components required for laparoscopic AR (our accompanying use case), show existing academic systems and address remaining challenges.

Our proposed AVB framework was inspired by the fact that virtual simulators in other fields have gained wide acceptance for evaluating novel AR visualizations. Medenica et al. [10] compare an experimental AR navigation system with two established navigation systems in a VR vehicle simulator, which provides a consistent environment and allows repeatable experiments. Tiefenbacher et al. [13] propose an architecture for the evaluation of mobile AR applications in VR. A Cave Automatic Virtual Environment (CAVE) was used to provide a reliable and reproducible virtual environment (industrial line) for the assessment of factors influencing the user experience.

Considering VR in medicine, Yiannakopoulou et al. [15] analyzed the effect of VR training and simulation for laparoscopic surgery and demonstrate their possibilities and limitations. Pfeiffer et al. [12] proposed the *IMHOTEP* VR framework (using the Unity game engine) for surgical applications, in which diverse preoperative patient data can be explored with the help of VR glasses. AVB combines the concept of virtual simulator skill training and VR environments in a unified evaluation framework, targeting surgical AR applications.

Methods

Objectively, evaluating the applicability and usability of visualization techniques in an AR setting for MIS is a challenging task. Live assessing the applicability of such methods poses a certain risk factor as visualizations might distract the surgeon from the focus region or, in the worst case, provide false information. Means to evaluate a method beforehand would reduce these influencing factors and improve the confidence of users [4]. To tackle this problem, we propose a steerable and interactive visualization evaluation framework, referred to as AVB, that is capable of virtually replicating intraoperative scenes, displaying AR in VR. This enables a comparison and exploration of parameter spaces of existing and new visualization techniques. More importantly, it provides physicians with a standardized baseline for evaluating visualization techniques.

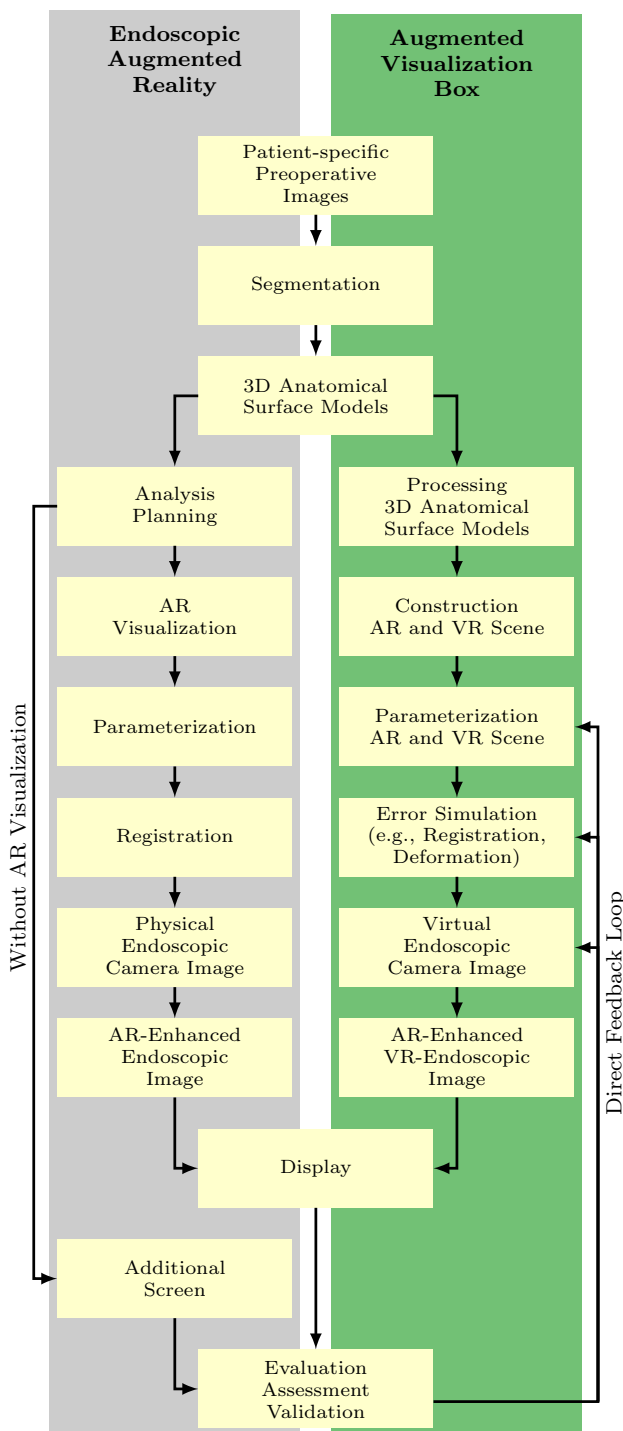


Fig. 1 Workflow of a conventional laparoscopic AR system (gray) compared to our proposed AVB approach (green). AVBs steps are structured according to the conventional procedure. The direct feedback loop during evaluation to the displayed VR scene allows parameter spaces to be examined

Laparoscopic augmented reality

Figure 1 provides an overview of high-level steps for laparoscopic AR. In short, patient-specific preoperative image

information is commonly acquired first. Then, relevant structures are segmented (e.g., blood vessels) to obtain 3D anatomical surface models. Depending on the surgical procedure, different analyses are performed, e.g., defining resection margins or a particular access path. Without AR, these preoperative data are displayed on an additional screen in the operating room (OR) and viewed on demand during surgery. After (preoperative) parameterization and (potentially successful) registration the AR visualization is superimposed on the physical endoscopic camera stream and, finally, displayed to surgeons.

Augmented Visualization Box

AVB provides a standardized baseline for dynamic environments and allows physicians to inspect and interact with preoperative planning data in VR while simultaneously viewing different AR visualization methods. The first three steps in the workflow of AVB are the same as in the conventional workflow (Fig. 1). The preoperative planning data are used, more precisely the 3D anatomical surface models, to create a virtual surgical scenario. In order to resemble the clinical setting as close as possible, AVB keeps two scenes simultaneously: The VR scene simulates the physical intraoperative endoscopic image and the AR scene is comprised of the superimposed content.

Processing of anatomical surface models

The preoperatively obtained surface models have to be augmented with texture information for advanced rendering. To accomplish this, surface normals and texture coordinates are computed and optimized to provide a uniform texture mapping. Organ-specific textures further enhance the impression of the VR scene, as shown in Fig. 2.

Construction of the VR and AR scene

After processing, the surface models have to be assembled to form a 3D scene (e.g., abdominal cavity). The appearance of the surfaces are designed using physically-based graphics shaders. Surgical instruments as well as a virtual endoscope are created, placed at their access ports and linked with application inputs, which allows interaction and manipulation at runtime. Simple keyboard and mouse input control events (e.g., switching influencing factors), while complex input devices such as joysticks or clinically used tracking devices interact directly with the scene or rather manipulate the virtual surgical instruments.

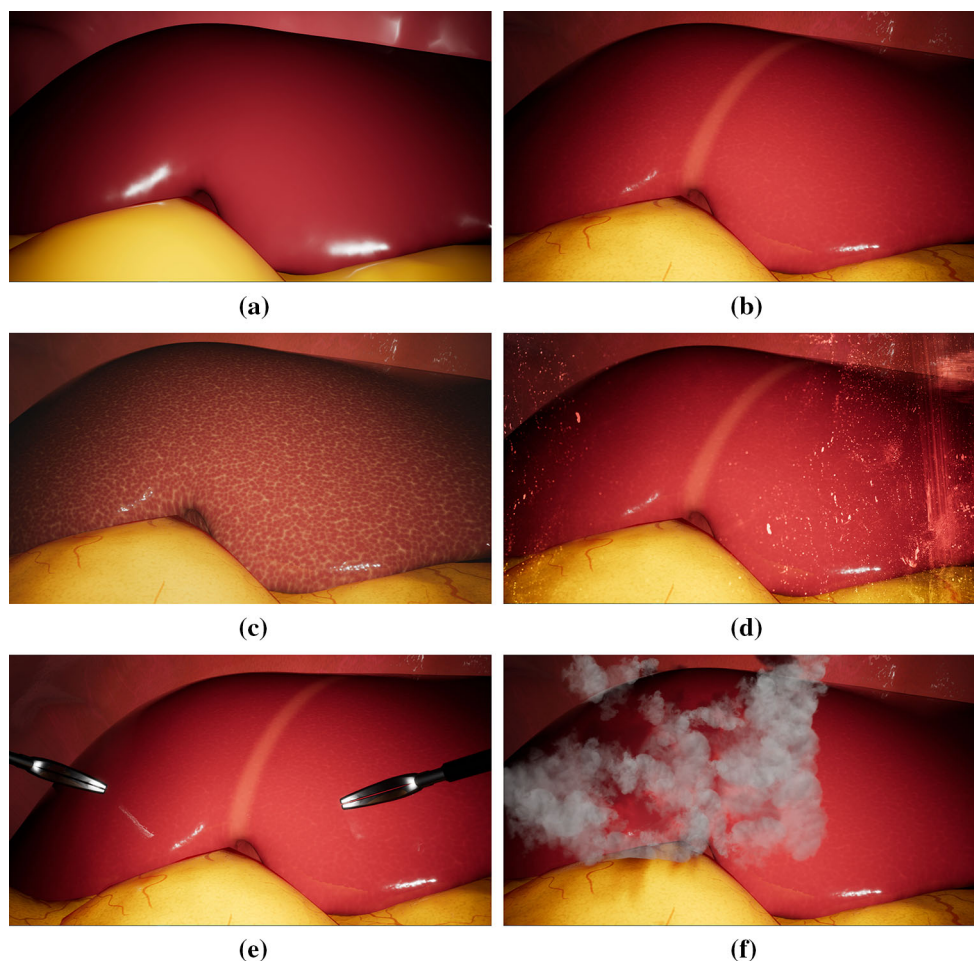


Fig. 2 Laparoscopic VR scene under different influencing factors: **a** simple shading (no texture), **b** advanced shading with textures (baseline), **c** alternative liver texture, **d** contaminated endoscope optics, **e** additional surgical grasping instruments and **f** simulated smoke

Parameterization of VR and AR scene

AVB is able to customize the appearance of both scenes as well as the position and orientation of the anatomical structures. As the structures for the AR and the VR scene stem from the same data sets, tedious registration procedures are superfluous. On the other hand, the settings can generate a controlled registration error to further analyze the influences of such inaccuracies.

Table 1 lists several influencing factors that can be adjusted individually or in combination at runtime. The *clinical factors* define characteristics that can be adapted before and during a surgery. *Design factors* describe the developer's ability to influence the VR scene in order to depict intraoperative uncontrollable conditions. One of the benefits of the AVB's architecture is a *direct feedback loop* that connects the evaluation, assessment and validation step backwards to any prior steps.

Implementation

We implemented AVB using the Unreal Engine 4.16.¹ This computer games engine (CGE) offers a development environment that allows users to rapidly prototype interactive photorealistic applications. For our application, the CGE provides a real-time global lighting model (adjustment of intraoperative lighting conditions), a flexible material editor (appearance of anatomical surfaces) and an editor for particle systems (Table 1). In addition, many post-processing options are provided to prototype different visualizations for AR.

We used the image processing software MeVisLab² to generate the 3D anatomical surface models. Afterward, 3D Studio Max 2018 was utilized to process the 3D surface models, to generate texture coordinates and to create 2D textures. To design an authentic VR scene, we created surface models of the laparoscopic grasping instruments and the trocar.

¹ <https://www.unrealengine.com/>.

² <https://www.mevislab.de/>.

Table 1 Real-time adjustable influencing factors of the VR scene

Description	Value
Clinical	
Light intensity	Off up to overexposed
Camera angle	0°, 30° and 45°
Camera FOV	30°–120°
Camera position	XYZ (mm)
Camera orientation	α , β , and γ (°)
Lense contamination	Off and on (scalable)
Depth of field	Focus point
Design	
Organ appearance	Colored or textured
Organ shape	Deformation
Organ position	XYZ (mm)
Organ rotation	α , β , and γ (°)
Smoke	Off and on
Lense distortion	Off and on
Light color	RGB

Clinical factors can be changed during surgery, whereas design factors influence the appearance of the VR scene, but are consistent during surgery

The user interaction was realized using a settings menu with mouse control, keyboard input and a G-Coder Simball 4D joystick³ that controls the surgical instruments.

The shader applied to the liver surface, abdominal wall and surrounding fat uses subsurface scattering. The surface textures are modified using procedurally generated roughness and reflection maps to simulate details such as veins or moisture. This is done by calculating roughness and reflection textures for the liver and soft tissue structure using Voronoi noise. Subsequently, those textures are further linearly interpolated with either manually designed vessel textures and simple Gaussian noise to obtain an even, natural surface finish. The color of the tissues can be adjusted applying an RGB color blending on the base texture. For an exceptional liver surface (cirrhosis), a manual design of the textures is recommended.

We applied two common visualization concepts to the AR scene found in the literature [8]: transparency and contour rendering (see Fig. 3). Both visualizations can be switched on and off. Additionally, the opacity of the overlay can be adjusted. For the contour visualization, post-processing methods are applied to calculate the contours using edge detection of the scene stored in a frame buffer. Therefore, the neighborhood of each pixel is used to examine discontinuities based on the depth or normal information in the frame buffer to specify edges. The color of the contours can also be specified by the users.

³ <http://g-coder.com/>.

For the implemented laparoscopic VR and AR scene, the liver including portal vein, hepatic vein and hepatic artery, 19 tumor metastases, as well as the surrounding organs with fat were segmented from computed tomography (CT) data by an expert. 3D anatomical surface models were created and further processed as described above. Textures were designed under supervision of two anatomical scientists to represent distinctive anatomical landmarks, e.g., falciform ligament (baseline). An alternative liver texture (see Fig. 2c) was designed darker in color than the baseline texture, which has more noticeable fat structures and no visible falciform ligament. The VR scene comprises the anatomical structures visible in the virtual endoscopic camera image, such as the liver, abdominal wall and fat. Each VR scene was rendered with a virtual endoscope having a 30° optics, as conventionally used in laparoscopy. The AR scene includes all visually obstructed anatomical structures, such as blood vessels and tumors. This scene is displayed as one of the two described AR visualizations, as shown in Fig. 3. The final scene is composed by the AR scene superimposed onto the VR scene (see Fig. 4).

To add a new AR visualization to AVB, the material or post-processing of the AR scene has to be extended. For this purpose, a new shader is created using the Unreal Engine's material editor that allows accessing various input data such as depth, color or position information. The engine creates a HLSL shader for the graphical representation of the AR visualization based on the created material.

Evaluation

In this section, the conducted study is detailed that evaluates the AVB with eleven surgeons by a dedicated questionnaire concerning (1) the VR scene, (2) the AR visualization and (3) necessary quantitative information for an AR visualization method.

We recorded the participants' demographic data and level of experience. Our study consisted of eleven participants, including 3 senior, 2 consultant and 4 resident surgeons and 2 medical students with surgical experience. They were on average 32.6 years old (range: 24–56), had an average of 5.4 years of professional experience (range: 0.5–20), 6 were female and 5 were male. Six participants claimed visual impairment (nearsightedness).

In the first part, the perception of the presented VR scene was assessed regarding visual realism. The participants were asked to evaluate the virtual endoscopic camera image (see Fig. 2) according to the following five established criteria described by Elhelw et al. [5]: specular highlights (light reflections), 2D texture detail (color and pattern on surfaces), 3D surface details (geometrical shape of organs and tissue), silhouettes (edges and outlines of geometric shapes) and depth visibility (perceptible depth cues). To explain the

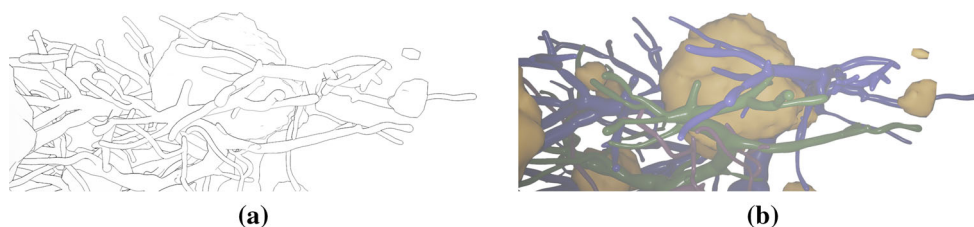


Fig. 3 Exemplar AR visualizations showing the delicate internal anatomical structures of the liver: contour (a) and transparency (b) visualization

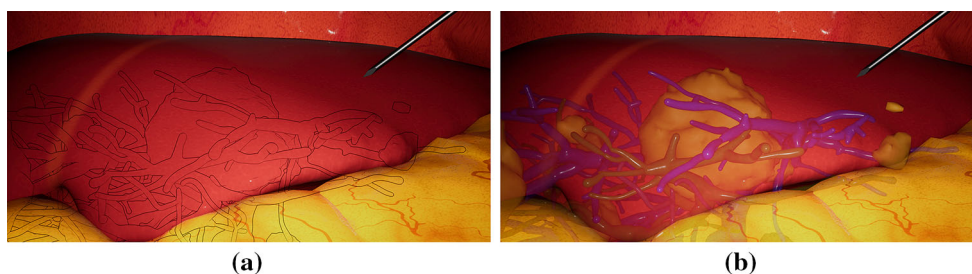


Fig. 4 Laparoscopic VR scene with an inserted needle enhanced with two AR superimpositions: contour (a) and transparency (b) AR visualization

criteria, corresponding definitions were given before the evaluation. All participants were asked to compare the presented series of six monoscopic video sequences (see Fig. 2) based on their professional experience in laparoscopic surgeries. We explicitly decided to restrict the study to generated video sequences instead of allowing direct interaction with the VR environment. Each sequence mimics a clinically realistic motion that consisted of movement toward and alongside the liver, similarly as described in [14]. Moreover, the order of the presented sequences was randomized. Additional comments during the assessment were collected.

Within the second part of our study, we assessed the applicability of AVB for the evaluation of AR in VR with regard to a defined task. A virtual ablation needle was placed in the VR scene. The task for each participant was to decide whether the displayed ablation needle will hit or miss a specific target structure. This decision making was supported with the described AR visualizations. As the instrument was positioned two to five centimeters away from the target location, the participants had to mentally prolongate the instrument in its pointing direction to make their decisions. Additionally, all participants had to report the confidence of their decisions. We generated three different needle positions and displayed each with the two AR visualizations on the baseline VR scene (see Fig. 3). As before, the order of needle position and visualization method was randomized.

Finally, we asked the participants to give feedback on quantitative and analytical information that should be conveyed by an AR visualization to maximize decision support during surgery. Furthermore, the experts informally commented on important image details, areas and relevant anatomical structures.

Results

VR evaluation outcomes

Table 2 lists the VR scene evaluation results with regard to visual realism and importance of the five criteria described above. Importance describes thereby a subjective rating of the physician, which gives a statement about necessary visual quality of the virtual image.

The **specular highlights** in the virtual environment have been judged most realistic for the baseline (3.32 ± 0.95) and the alternative liver texture (3.27 ± 1.01). The other scenes were valued slightly weaker.

The evaluation revealed that **2D texture details** have a very high impact on the visual perception of the VR scene. The video sequence showing the alternative liver texture (3.91 ± 0.83) was rated higher than the baseline (3.32 ± 0.84). In contrast to this, the scene without appropriate surface textures was rated to have lowest realism (1.91 ± 0.83) because of its artificial appearance. In comparison with that, instruments and contamination had little influence on the visual perception of the 2D texture details.

The **3D surface details** achieved similar results as the 2D texture details. The videos for baseline as well as the alternative liver texture had the best results with 3.55 ± 0.96 and 3.82 ± 0.87 , respectively. The untextured anatomical surfaces and the simulated smoke reduce the overall impression of realism.

Silhouettes can greatly impair the visual quality if they appear artificial. The VR scene without textures emphasizes non-organic (angular) organ borders (2.91 ± 1.22) that result from segmentation errors and mesh smoothing. Surface textures (3.73 ± 0.90) or influencing factors such as lens con-

Table 2 Mean values and standard deviation assessing the visual realism and importance of the VR scene

Scene	Specular highlights	2D texture details	3D surface details	Silhouettes	Depth visibility
Visual realism					
Baseline	3.32 ± 0.95	3.32 ± 0.84	3.55 ± 0.96	3.23 ± 1.11	3.55 ± 0.96
Liver texture	3.27 ± 1.01	3.91 ± 0.83	3.82 ± 0.87	3.73 ± 0.90	3.45 ± 1.04
Without texture	2.91 ± 0.94	1.91 ± 0.83	2.91 ± 0.94	2.91 ± 1.22	2.91 ± 0.70
Instruments	2.91 ± 1.22	3.27 ± 0.65	3.45 ± 1.04	3.45 ± 1.04	3.18 ± 1.25
Contamination	2.91 ± 1.14	3.18 ± 0.75	3.45 ± 1.21	3.27 ± 0.79	3.09 ± 0.94
Smoke	2.55 ± 0.82	2.73 ± 0.79	2.64 ± 0.81	3.09 ± 1.14	2.86 ± 1.00
Importance					
Baseline	3.73 ± 0.70	4.36 ± 0.73	4.55 ± 0.49	4.36 ± 0.58	4.55 ± 0.67
Liver texture	3.82 ± 0.60	4.45 ± 0.52	4.41 ± 0.49	4.36 ± 0.67	4.45 ± 0.69
Without texture	3.91 ± 0.70	4.36 ± 0.67	4.50 ± 0.50	4.27 ± 0.47	4.55 ± 0.52
Instruments	4.00 ± 0.77	4.45 ± 0.52	4.45 ± 0.69	4.27 ± 0.65	4.55 ± 0.69
Contamination	3.91 ± 0.94	4.36 ± 0.67	4.59 ± 0.49	4.27 ± 0.47	4.45 ± 0.69
Smoke	4.00 ± 0.77	4.27 ± 0.47	4.36 ± 0.67	4.45 ± 0.69	4.55 ± 0.69

Results are based on a 5-point Likert scale from 1 = not realistic/unimportant to 5 = realistic/important

tamination (3.27 ± 0.79) or even instruments (3.45 ± 1.04) visually obstruct these details, therefore they are less prominent in the other scenes.

The importance of **depth perception** was found to be relatively constant across all presented video sequences. The baseline and alternative liver texture showed the best results (3.55 ± 0.96 and 3.45 ± 1.04). The smoke had the greatest influence on depth perception (2.86 ± 1.00). Moreover, the importance across all criteria remains almost unchanged.

Individual comments and decisions made in the questionnaire were discussed with the participants. Overall, we received positive feedback, but the following points were addressed by the participants. During the assessment of the VR scene, the depiction of the smoke was criticized because it spreads differently in real scenarios and might obscure the image completely. Furthermore, the reflection on the surgical instruments was too weak or the image was too dark. The contamination of the lens was too sharply perceived by four participants. In a real scenario, the endoscope image blurs due to different focus points if the optics are smudged by fluids. Another issue was the resolution of the surface models, which should be smoother for a more natural appearance.

AR in VR evaluation outcomes

We further assessed, whether the surgeon was able to make a correct decision based on the AR visualization before the visualization is used in the OR. This means that a surgeon is able to make a correct decision, i.e., to reliably target a specific structure using a surgical instrument, with the help of an AR visualization. Reliable or rather correct in this context means that it is possible to safely estimate whether a needle

Table 3 Results of the second part of the study assessing a targeting task with two AR visualization in VR

AR visualization	Needle position	Correct	Incorrect	Confidence
Transparency	1	5	6	2.45 ± 0.88
	2	7	4	2.64 ± 1.09
	3	1	10	3.55 ± 1.33
Contour	1	5	6	2.45 ± 1.01
	2	6	5	2.64 ± 1.09
	3	4	7	3.09 ± 1.32

Absolute values of correct as well as incorrect answers. Mean value and standard deviation of the corresponding confidence for three needle trajectories (rated on a 5-point Likert scale from 1 = unsure to 5 = sure)

will hit the tumor. The results listed in Table 3 apparently show that it is challenging to make a reliable judgment of targeting tasks with a simple overlay visualization. Participants erroneously decided that the indicated needle would hit the tumor even though it did not, and vice versa, namely 61% for the transparency visualization and 55% for the contours. Ten out of eleven (91%) participants made a false statement whether needle 3 will hit the tumor or not when using the transparency visualization, but were very confident that their judgment was correct. When using the contour visualization, seven out of eleven (63%) participants were mistaken, but were very confident that they made the right decision. Needle 3 was located at a distance of about 2.5 cm from the liver surface and perpendicular to the viewing direction. The minimum distance of the prolonged needle path to the tumor was 2 and 5 cm to the tumor center. The participants stated that the tumor size led to the assumption that it is hit by the

needle. The results for needle 1 and 2 are equally weighted both in terms of correctness and confidence. For needle 1, 55% of the participants were wrong in either case. In case of needle 2, the false assessment rate for the transparency was 37% and for the contours 45%. However, compared to needle 3, the participants were less confident in their decision. This shows that a supporting AR visualization may increase the surgeon's confidence in making clinical decisions, but at the same time enforce erroneous decisions.

According to the surgeons, the most important information for conducting minimally invasive surgery with AR support is:

- positions and spacings of anatomical structures and surgical instruments (10 of 11),
- distances and relationships between the structures, e.g., which structures lie behind another structure (9 of 11) and
- depth relations (10 of 11).

Absolute values such as tumor volume, area or task and planning completion time are less relevant for orientation, but might provide useful information for other clinical tasks (e.g., resection) in which planning is important.

Discussions

The AVB offers many possibilities to actively influence the VR scene. Starting from clinical factors such as light intensity or endoscope angle up to different design factors, e.g., appearance of anatomical surfaces. Thus, it is not only possible to examine a scene with a newly developed visualization, but to examine several variations of it. Furthermore, influencing factors can be simulated which are difficult to investigate systematically in an intraoperative scenario.

The conducted survey has pointed out that the 2D surface texture in particular have a high influence on the visual realism. Furthermore, the evaluation of the simple AR visualizations has shown that without context or references it is difficult to make reliable decisions. Even if the task the surgeons had to decide on was very simple, they mostly failed. This motivates the development and standardized assessment of more sophisticated visualization method, that support the surgeon when conducting the task. The number of participants was relatively small (eleven) and some of them had limited professional experiences. Further investigations with a larger and more experienced number of surgeons (e.g., senior surgeons) should be carried out in the future.

Another advantage of AVB is its transfer to other intraoperative scenarios, e.g., endoscopic kidney surgery. Further applications with new 3D surface models and internal structures of the liver or kidney can be evaluated directly without having to design and set up a completely new system.

One major benefit of AVB is that the registration of the AR scene to the VR scene is 100% accurate, not considering any potential errors from the preceding segmentation or surface model generation steps. Hence, the laboratory conditions of an evaluation can be preserved for assessing visualization techniques across multiple sessions or different methods according to a standardized baseline within a single session. However, this possibility can also be used to investigate the effects of registration errors in a controlled manner by changing the position and orientation of the AR scene.

Conclusion

In this paper, we presented the concept of the Augmented Visualization Box (AVB), that offers the possibility to examine the AR visualizations in VR in an objective and standardized way. AVB was investigated in a user study with eleven surgeons on its applicability and adaptability, with special focus on the perceived visual realism of the VR scene and the assessment of AR visualizations in VR. The main benefit of AVB is that the AR visualization method can be assessed solely according to perceptual factors, isolating this important step from other potential system inaccuracies that could be a source of procedural failure such as registration errors. It is therefore possible to control various influencing factors in order to better explore larger parameter spaces of new AR visualization methods. Our results show that surgeons have more confidence in their decisions when supported by AR visualizations, although they are mostly wrong. As potential avenue for future work, this outcome has to be investigated in order to translate AR visualizations into the OR.

In the future, further examinations will be carried out in order to improve visualizations and prepare them for intraoperative use. Furthermore, the influence of stereo laparoscopy has to be investigated in consequence of the technical progress and broader availability in hospitals. Due to the stereoscopic depth cues, new challenges for visualizations could arise. Finally, the effects of organ deformations should be considered. For this purpose, AVB has to be extended with a physics simulation that can be applied in a controlled way on the VR and AR scene to examine deformation influences.

Acknowledgements We thank Fraunhofer MEVIS, Bremen, Germany for providing segmentation masks of the liver, tumors and vascular structures with which we have created our 3D models of the liver.

Compliance with ethical standards

Funding This work is partially funded by the Federal Ministry of Education and Research (BMBF) within then STIMULATE research campus (Grant Number 13GW0095A), by the European Regional Development Fund under the operation number 'ZS/2016/04/78123'

as part of the initiative “Sachsen-Anhalt WISSENSCHAFT Schwerpunkte” and by the German Research Foundation (DFG) project EN 1197/2-1.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standard For this type of study, formal consent is not required.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

1. Amir-Khalili A, Nosrati MS, Peyrat JM, Hamarneh G, Abugharbieh R (2013) Uncertainty-encoded augmented reality for robot-assisted partial nephrectomy: a phantom study. In: *Augmented reality environments for medical imaging and computer-assisted interventions*, Springer, pp 182–191
2. Bernhardt S, Nicolau SA, Soler L, Doignon C (2017) The status of augmented reality in laparoscopic surgery as of 2016. *Med Image Anal* 37:66–90
3. Bichlmeier C, Sielhorst T, Heining SM, Navab N (2007) Improving depth perception in medical ar. In: *Bildverarbeitung für die Medizin 2007*, Springer, pp 217–221
4. Dixon BJ, Daly MJ, Chan H, Vescan AD, Witterick IJ, Irish JC (2013) Surgeons blinded by enhanced navigation: the effect of augmented reality on attention. *Surg Endosc* 27(2):454–461
5. Elhelw M, Nicolaou M, Chung A, Yang GZ, Atkins MS (2008) A gaze-based study for investigating the perception of visual realism in simulated scenes. *ACM Trans Appl Percept* 5(1):3:1–3:20
6. Hansen C, Wiefelich J, Ritter F, Rieder C, Peitgen HO (2010) Illustrative visualization of 3d planning models for augmented reality in liver surgery. *Int J Comput Assist Radiol Surg* 5(2):133–141
7. Katić D, Wekerle AL, Görtler J, Spengler P, Bodenstedt S, Röhl S, Suwelack S, Kenngott HG, Wagner M, Müller-Stich BP, Dillmann R, Speidel S (2013) Context-aware augmented reality in laparoscopic surgery. *Comput Med Imaging Graph* 37(2):174–182
8. Kersten-Oertel M, Jannin P, Collins DL (2013) The state of the art of visualization in mixed reality image guided surgery. *Comput Med Imaging Graph* 37(2):98–112 (**special Issue on Mixed Reality Guidance of Therapy - Towards Clinical Implementation**)
9. Lerotic M, Chung AJ, Mylonas G, Yang GZ (2007) pq-space based non-photorealistic rendering for augmented reality. In: Ayache N, Ourselin S, Maeder A (eds) *Medical image computing and computer-assisted intervention—MICCAI 2007*, pp 102–109
10. Medenica Z, Kun AL, Paek T, Palinko O (2011) Augmented reality versus street views: a driving simulator study comparing two emerging navigation aids. In: *Proceedings of the 13th international conference on human computer interaction with mobile devices and services, ACM, New York, NY, USA, MobileHCI '11*, pp 265–274
11. Nicolaou M, James A, Lo BPL, Darzi A, Yang GZ (2005) Invisible shadow for navigation and planning in minimal invasive surgery. In: Duncan JS, Gerig G (eds) *Medical image computing and computer-assisted intervention—MICCAI 2005*, pp 25–32
12. Pfeiffer M, Kenngott H, Preukschas A, Huber M, Bettscheider L, Müller-Stich B, Speidel S (2018) Imhotep: virtual reality framework for surgical applications. *Int J Comput Assist Radiol Surg* 13:741–748
13. Tiefenbacher P, Lehment NH, Rigoll G (2014) *Augmented reality evaluation: a concept utilizing virtual reality*. Springer International Publishing, Cham, pp 226–236
14. Wild E, Teber D, Schmid D, Simpfendorfer T, Müller M, Baranski AC, Kenngott H, Kopka K, Maier-Hein L (2016) Robust augmented reality guidance with fluorescent markers in laparoscopic surgery. *Int J Comput Assist Radiol Surg* 11(6):899–907
15. Yiannakopoulou E, Nikiteas N, Perrea D, Tsigris C (2015) Virtual reality simulators and training in laparoscopic surgery. *Int J Surg* 13:60–64